Taylor, D.W. 2014. Large inequalities in herbarium specimen density in the western United States. Phytoneuron 2014-53: 1–8. Published 2 June 2014. ISSN 2153 733X

# LARGE INEQUALITIES IN HERBARIUM SPECIMEN DENSITY IN THE WESTERN UNITED STATES

**Dean Wm. Taylor**
Redwood Drive
Aptos, California 95003-2517
deanwmtaylor@gmail.com

## ABSTRACT

Herbarium specimen records from 11 western states of the USA are assayed (5.6 million specimen records obtained from on-line databases). On a county basis, specimen density varies considerably, by a factor of ~2500 (highest in San Francisco County, California, at 99.2 specimens/km$^2$; lowest in Blaine, Liberty, and Prairie counties, Montana, at 0.04 specimens/km$^2$). California exhibits the highest specimen density, nearly 2x higher than the second highest state (Wyoming). All states except Nevada have one to several counties which fall above the 80$^{th}$ percentile for density. Many states have large areas below the 20$^{th}$ percentile. Specimen density is correlated with human population density. Large areas deficient in specimen records indicate challenges for vascular plant systematics and conservation. A comprehensive, adequately funded program of floristic exploration in the western USA would require gathering of an additional ~1.5-2 million specimens to minimally rectify the observed inequalities.

Herbarium specimens are the fundamental currency of biodiversity studies. In the intellectual economy of biological systematics, classifications are products, specimens their raw materials. In the USA, herbaria have made great progress in digitization of specimen records in their holdings, supported by modest funding. Production of regional floras such as the revised *Jepson Manual* (Baldwin et al. 2012) and Flora North America North of Mexico (FNA 1993-2006) have been greatly facilitated by digitization of herbarium records. However, the increased access to digitized specimen records is in contrast to the overall decline in specimen collecting (Prather et al. 2004a,b; Tewksbury et al. 2014).

This paper identifies a significant inequality of herbarium specimen density in the western USA, based on accession of digitized records.

## Methods

I queried on-line specimen databases for herbarium records of vascular plants by county for the 414 counties in 11 western USA states between 10 January and April 4, 2014. Records from regional consortia were tallied first (Consortium of California Herbaria, Consortium of Pacific Northwest Herbaria, Southwest Environmental Information Network), then records from larger individual herbaria were added (CAS, RM, COLO, MINN and MO-Tropicos — herbarium abbreviations follow Thiers 2014). Queries were structured to eliminate specimen records offered by more than one data provider across platforms. Data on spelling of county names, state and county land surface areas, and human population density (2010 census) were obtained from Wikipedia (accessed January 2014). Several variant county spellings were corrected in the queries. Minor internal inconsistencies in the area of counties and states were not corrected (these in no case are greater than about 20 km$^2$). Presentation and analysis of the data is given as percentiles for simplicity of comparisons.

**Results and discussion**

Observed specimen density (specimens/km$^2$) differs by a factor of 2480, ranging between 99.2 (San Francisco County, California) to 0.04 (Blaine, Liberty, and Prairie counties, Montana). Using the breakdown criteria of Tukey (1977), 33 counties are outliers relative to the other 381 counties: all outliers are enriched in specimen density. The mean specimen density of these 33 enriched outliers is 12.27 specimens/km$^2$. The mean for the 381 counties that are not outliers is 1.65 specimens/km$^2$, smaller by a factor of 7. Figure 1 provides a frequency distribution of the density for the 414 counties: the distribution is highly left-skewed.

The inequalities in specimen density have a strong geographic pattern (Figure 2, Table 1). Relative to other states, California is well-collected, with an average density of 4.76 specimens/km$^2$ (1.9 million digitized specimen records). Allowing for the fact that California herbaria are not yet fully digitized (CAS and DAV remain to finish), the actual density might approach 6 specimens/km$^2$ (based on reported *Consortium of California Herbaria* digitization progress reports). By contrast, Montana and the Great Basin states of Nevada and Utah are comparably depauperate (Table 1).

All states have both well-collected (defined as >80$^{th}$ percentile) and under-collected <defined as <50$^{th}$ percentile) counties (Figure 2). Some of the deficient counties are remote from population centers (Utah, Carbon, and Emery counties; Nevada, all but Carson City; Crook County, Oregon, etc., cf. Figure 2). Other deficient regions include counties with a large fraction of the land area devoted to agriculture (e.g. Jerome and Minidoka counties, Idaho; counties in eastern Colorado). Other deficient counties, although agricultural, include sizeable remnants of non-managed vegetation (most of the deficient counties in Montana; Jefferson County in Idaho; Adams, Franklin, and Lincoln counties in Washington). The frequent low specimen density in rural counties also suggest that weed records might be disproportionately lacking therein.

Regression analysis identified a significant correlation between specimen density and density of human population based on the 2010 census ($R^2$ 0.66, $p<0.001$) — counties with a greater human population are well-collected compared to sparsely populated counties. Moreover, the geographic location of enriched vs. deficient counties suggests that proximity to herbaria and universities, which are located in population centers, is important. Such a pattern is not that unexpected: botanists gather specimens close to home.

The western USA states fall perhaps into three groups (Table 1). California, Colorado, and Wyoming have high to moderate specimen density, Washington and Arizona intermediate density, with the remaining states trailing behind (Nevada far in the rear).

Table 1 summarizes two potential criteria upon which the adequacy of regional specimen density could be assessed. If states are judged relative to a performance standard based on the mean specimen density for California, nearly 9 million additional specimen records would be necessary to attain this data density elsewhere over the region. Alternatively, if states are judged against the 70$^{th}$ percentile for non-California counties (about 2 specimens/km$^2$), the resultant specimen deficit is a mere 1.8 million specimens (Table 1). By this latter standard California, Wyoming, and Colorado are well-collected, with Washington, Arizona, Idaho, Oregon, New Mexico, Utah, Montana, and Nevada deficient to varying degrees.

For convenience, readers who know the specimen number or density for a particular county can approximate percentile using the regression equation: percentile = 0.2246ln (density) + 0.4251 ($r^2$=.939)

| | Specimen Density (sheets/km2) | | | | | Deficit Comparison | |
|---|---|---|---|---|---|---|---|
| STATE | Mean | High | Low | Specimens | Area ($km^2$) | Calif. Standard | 70% tile Non-CA Standard |
| California | **4.76** | 99.19 | 0.35 | 1,922,851 | 404,224 | 0 | +1,118,445 |
| Wyoming | **2.75** | 5.82 | 0.83 | 695,570 | 253,348 | -510,366 | +191,407 |
| Colorado | **2.30** | 14.74 | 0.06 | 620,782 | 269,837 | -663,642 | +83,806 |
| Washington | **1.92** | 26.84 | 0.03 | 355,227 | 184,827 | -524,550 | -12,579 |
| Arizona | **1.84** | 6.21 | 0.67 | 542,300 | 295,234 | -863,014 | -45,216 |
| Idaho | **1.43** | 7.58 | 0.08 | 310,419 | 216,632 | -720,749 | -120,679 |
| Oregon | **1.34** | 10.73 | 0.36 | 340,518 | 255,026 | -873,406 | -166,984 |
| New Mexico | **1.01** | 10.98 | 0.05 | 317,580 | 315,194 | -1,182,743 | -309,656 |
| Utah | **0.82** | 3.76 | 0.16 | 179,546 | 219,887 | -867,116 | -258,029 |
| Montana | **0.81** | 4.86 | 0.04 | 310,181 | 380,800 | -1,502,427 | -447,611 |
| Nevada | **0.34** | 2.25 | 0.13 | 96,696 | 286,367 | -1,266,411 | -473,174 |
| Grand Mean | **1.77** | | | **5,691,670** | **3,081,376** | **-8,974,425** | **-1,833,927** |

Table 1. Number of databased specimens by state, specimen density, and deficit comparisons based on the specimen density for California, and for the 70$^{th}$-percentile county density for non-California counties.

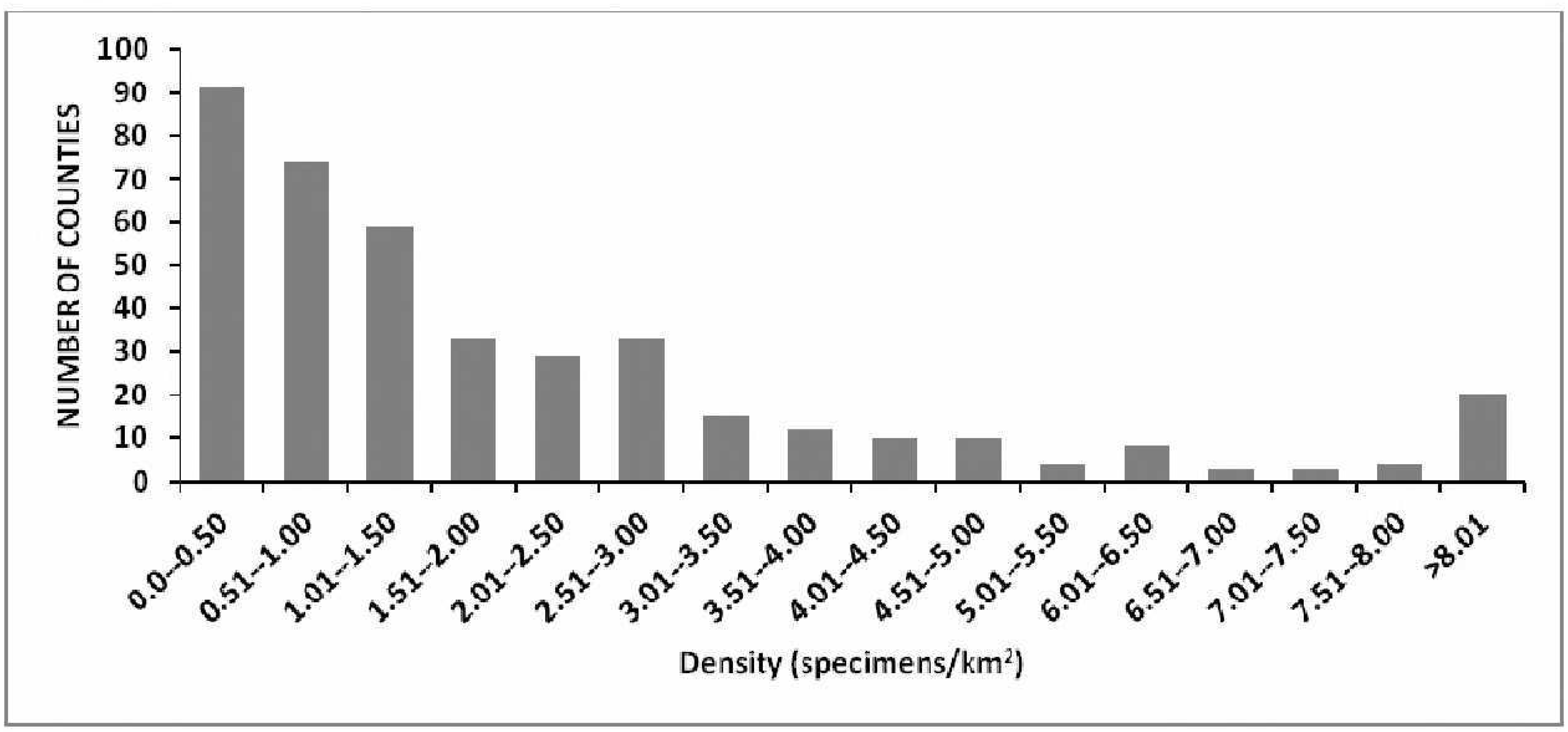


Figure 1. Frequency distribution of specimen density for 414 counties in 11 western states of the USA. The distribution is strongly left-skewed (the mean density is 1.77, Table 1).

**Arizona** – Only one county, the smallest in the state, Santa Cruz County, is well-collected, with adjacent Pima and Cochise counties intermediate. Notable in this region are the high mountains with Madrean floristic affinities (McLaughlin 1994), which are regarded as a biodiversity hot-spot (Mittermeier et al. 2011) and which are under active study (see Felger et al. 2014). Two of the deficient counties (Yuma and La Paz) are in the western Sonoran desert region: adjacent Imperial County, California, and Clark County, Nevada, have similarly low density, making a regional deficiency. Navajo County is likewise deficient — its far northern portion is included in the recent *Four Corners Flora* (Heil et al. 2013).

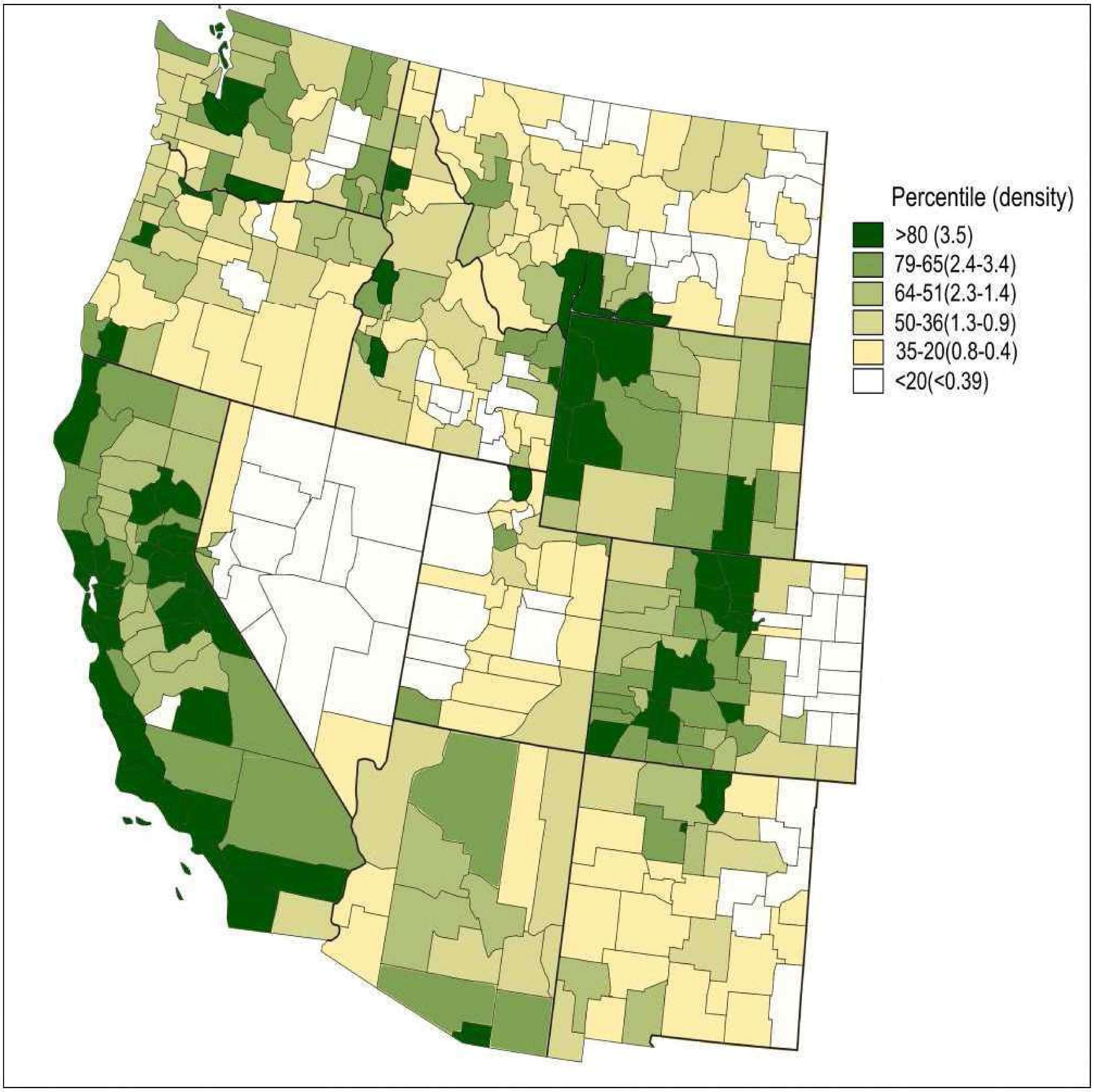


Figure 2. County map of the western United States showing herbarium specimen density (6 mapped categories based on percentile of observed county density variation).

**California** - In California, which is generally well-collected, Kings County stands out as one of the least collected counties in the western USA (0.32 specimens/km$^2$, 15$^{th}$ percentile). It would be attractive to attribute the lack of specimens from Kings County to the fact that a large portion of the county lies in the San Joaquin Valley, the most intensively agrarian landscape in North America (Preston 1981). However, the western third of the county is mountainous, situated in the inner South Coast Ranges, suggesting the lack of specimens is not due to a lack of species richness in the flora. The recently described monotypic Brassicaceae genus *Twisselmannia* (Al-Shebaz 1999), a narrow endemic to Kings County, suggests that the low collection density for that county may owe, at least in part, to under-collection.

The map in Figure 2 shows that California contains a disproportionate number of well-collected counties. Moreover, non-political geographic subdivisions of the state are also high density outliers: Yosemite National Park (33,214 specimens) has a density of 11 specimens/$km^2$, equal to $97^{th}$ percentile for counties. Yosemite Valley (±20 $km^2$, 5843 specimens) has an extreme density, about 275 specimens/$km^2$. The California Channel Islands are also all extreme high density outliers: Anacapa Island (911 specimens/$km^2$), Santa Barbara Island (488), Santa Cruz Island (87), Santa Catalina Island (60), San Nicolas Island, and Santa Rosa Island (28 specimens/$km^2$) all exceed the $99^{th}$ percentile equivalent for county density.

**Colorado** – third to California in collection density (Table 1), the strong pattern of well-collected counties in the mountains along the Front Range and in the headwaters Colorado River basin is in direct contrast to the poorly collected counties in the eastern Great Plains. Eleven counties exceed the $90^{th}$ percentile: Gilpin, Boulder, Clear Creek, San Juan, Grand, Denver, Lake, Larimer, Gunnison, Jackson, and Jefferson. Although the recently published *Four Corners Flora* (Heil et al. 2013) covers a geographic region where some of the counties are well-collected, overall the collection density of that flora region averages the $58^{th}$ percentile. The eastern plains portion of the state average in the $7^{th}$ percentile (in this region, Otero County is better collected, $51^{st}$ percentile). These eastern counties are largely agricultural and sparsely populated: low specimen density is opposed to the prevalence of grazing lands used for intensive livestock grazing, suggesting poor floristic documentation. The same west to east gradient in specimen density is also seen in adjacent New Mexico.

**Idaho** – overall collecting density is more uniform than adjacent states, but most counties fall in the lower range (<$50^{th}$ percentile). Overall, the density pattern is similar to Oregon, but there are still several marked inequalities. The low collection density in the western portion of the Snake River plains (Lincoln, Jefferson, Minidoka, and Jerome counties) indicates a significant regional deficit, perhaps attributable to an intensive agricultural district. The high concentration in Latah County ($85^{th}$ percentile) is part of a regional pattern connected to adjacent Washington (associated with proximity to WTU and nearby ID/IDF).

**New Mexico** – almost all of the state falls below the $35^{th}$ percentile, with the far eastern portion below the $20^{th}$ percentile (the easternmost deficient counties mirror the same pattern in Colorado). The sole well or moderately collected counties are in the north in the Rocky Mountains.

**Nevada** – relatively a 'third-world' state, the least collected state of any in the western USA. The recently finished *Intermountain Flora* (Holmgren et al. 2012) essentially covers most all the white space of the map in Figure 2. It should be noted that digitized specimen records for the *Intermountain Flora* region concentrate on the last two published volumes (Holmgren et al. 2005, 2012) and as such, on-line specimen density for the region is artificially low, although it is unlikely to be of notable greater overall magnitude. The only counties with even a modicum of specimen density are those in the Sierra Nevada and its eastern foothills adjacent to California (Carson City, Douglas. and Storey counties, in the Reno region).

**Montana** – herbarium specimen density for most of the state is dismally low, relatively. Half of the counties in the state are below the $20^{th}$ percentile, and nearly 90% of the state is below the $50^{th}$ percentile. Only three counties (Carbon, Gallatin, and Park) have high density (>$80^{th}$ percentile), these in the Rocky Mountains (proximal to MONT and more significantly RM) and adjacent to much better collected Wyoming. The large deficient counties are sparsely populated but are not entirely given over to intensive agriculture; they retain sizable remnant habitat areas of non-managed vegetation.

**Oregon** – specimen density is most distinctly divided regionally west and east of the Cascades, with the highest density in urban Multnomah and Benton counties (vicinity of ORE/OSU). The far southwestern counties (Josephine, Curry, and Jackson) are well-collected, being part of the California Floristic Province. The moderate density in the Wallowa Mountains of the far northeast is also of note.

**Utah** – significantly under-collected relative to the overall mean density (Table 1). The only two counties with high density (Daggett and Cache) are in the Wasatch Mountain province. Washington County in the far southwest is moderately well-collected, at 74$^{th}$ percentile. Most of the Bonneville Basin in the western half of the state is very significantly under-collected (<20$^{th}$ percentile). The state flora, now in its 4$^{th}$ revision (Welsh et al. 2008), indicates slow but steady progress in floristic documentation, but the low overall collection density suggests opportunities for fruitful future floristic exploration.

**Wyoming** – the second best collected state in the data set. The 6 counties that comprise the Rocky Mountain portion of the state – Teton, Albany, Park, Sublette, Lincoln, and Crook, form a large well-collected block exceeding the 80$^{th}$ percentile. By comparison to other western states, Wyoming has comparably few poorly collected counties: Niobrara County is the lowest at the 33$^{rd}$ percentile. Compared to ecologically similar Montana, the differences in collection density amongst comparable physiographic regions in the two states are manifold.

**Conclusions**

Large areas of the North American continent extremely deficient in herbarium specimen records indicate significant challenges for vascular plant systematics and conservation in the USA. Bebber et. al (2010, 2012) have quantitatively demonstrated that description of new species of vascular plants is disconnected from their initial discovery. That is, new species are most often "discovered" in herbaria — specimens of undescribed taxa reside in collections for several decades before they are recognized and treated and their initial recognition requires a threshold number of accumulated specimens. The nearly 2500-fold difference in collection density reported in this paper strongly implies that undiscovered plant diversity in under-collected regions is concealed. Given this, going forward, preparation of floristic treatments will suffer from the unequal distribution of specimens.

Bebber et al. (2012) also have shown that contribution of type specimens is non-randomly distributed among collectors. Their analysis of the dynamics of plant collecting shows that long apprentice times are required to develop expertise in floristic botany. Accordingly, the apparent decline in practicing taxonomists (Whitfield 2012; Bacher 2012) is in stark contrast to the large large, under-collected areas of Figure 2. Who will collect the requisite 2 million specimens to fill in the map?

The fact that California is disproportionately collected relative to the 10 other western USA states — arguably 'over-collected' — would seem to indicate that no further specimens need be gathered there. That such a view would be folly has been convincingly demonstrated by Joppa et al. (2011), who show that biodiversity 'hot-spots' house most undescribed plant species. The California region is such a 'hot-spot' of vascular plant diversity. In the roughly two decade interval between the 1$^{st}$ and 2$^{nd}$ editions of the *Jepson Manual* (Hickman 1993; Baldwin et al. 2012), about 163 new California endemics were described, and about 373,000 additional accessions were made after initial publication of a revised, modern flora (specimen data estimate based on *Consortium of California Herbaria*). Even well-collected California needs attention from floristic botanists.

Common sense indicates that it is unlikely that botanists can muster the effort, given available resources, to gather 9 million additional specimens from the western USA. I believe, however, that it is not unrealistic to gather and/or database an additional 1.8 million regional specimens (to equal the 70[th]-percentile standard, see Table 1). Attaining such a goal would require significant increase in floristic field studies, with increased and adequate funding. A comprehensive, adequately funded program of floristic exploration in the western USA is necessary to rectify the observed inequalities.

## LITERATURE CITED


Al-Shebaz, I. 1999. *Twisselmannia* (Brassicaceae), a remarkable new genus from California. Novon 9: 132–135.

Bacher, S. 2012. Still not enough taxonomists: Reply to Joppa et al. Trends Ecol. Evol. 26: 65–66.

Bebber, D.P., M.A. Carine, J.R.I Wood, A.H. Wortley, D.J. Harris, G.T. Prance, G. Davidse, J. Paige, T.D. Pennington, N.K.B. Robson, and R.W. Scotland. 2010. Herbaria are a major frontier for species discovery. Proc. National Academy Sciences U.S.A. 107: 22169–22171.

Bebber, D.P., M.A. Carine, G. Davidse, D.J. Harris, E.M. Haston, M.G. Penn, S. Cafferty, J.R.I. Wood, and R.W. Scotland. 2012. Big hitting collectors make massive and disproportionate contribution to the discovery of plant species. Proc. Royal Society B 279: 2269–2274.

Baldwin, B.G., D.H. Goldman, D.J. Keil, R. Patterson, and T.J. Rosatti (eds.). 2012. The Jepson Manual: Vascular Plants of California (ed. 2). Univ. of California Press, Berkeley

Felger, R.S., S. Rutman, and J. Malusa. 2014. Ajo Peak to Tinajas Altas: A flora of southwestern Arizona. Part 6. Poaceae – grass family. Phytoneuron 2014-35: 1–139.

Flora of North America Editorial Committee (eds.). 1993-2006. Flora of North America North of Mexico. 16+ vols. New York and Oxford.

Heil, K.D., S. O'Kane Jr., L.M. Reeves, and A. Clifford. 2013. Flora of the Four Corners Region: Vascular Plants of the San Juan River Drainage: Arizona, Colorado, New Mexico, and Utah. Missouri Botanical Garden, St. Louis.

Hickman, J.C. (ed.) 1993. The Jepson Manual, Higher Plants of California. Univ. of California Press, Berkeley.

Holmgren, N., P.K. Holmgren, and A. Cronquist. 2005. Intermountain Flora: Vascular Plants of the Intermountain West, U.S.A. Volume 2, Part B, Subclasses Dilleniidae. New York Botanical Garden, Bronx.

Holmgren, N., P.K. Holmgren and J.L. Reveal. 2012. Intermountain Flora: Vascular Plants of the Intermountain West, U.S.A. Volume 2, Part A, Subclasses Magnoliidae-Caryophyllidae. New York Botanical Garden, Bronx.

Joppa, L. N., D.L. Roberts, N. Meyers, and S.L. Pimm. 2011. Biodiversity hotspots house most undiscovered plant species. Proc. National Academy Sciences U.S.A. 108: 13171–13176.

McLaughlin, S.P. 1994. An Overview of the Flora of the Sky Islands, Southeastern Arizona: Diversity, Affinities, and Insularity, Biodiversity and Management of the Madrean Archipelago. The Sky Islands of Southwestern United States and Northwestern Mexico. U.S.D.A. Forest Service, GTR-264, Fort Collins, Colorado.

Mittermeier, R.A., W.R. Turner, F.W. Larsen, T.M. Brooks, and C. Gascon. 2011. Global biodiversity conservation: The critical role of hotspots. Pp. 3–16 in F.E. Zachos and J.C. Habel (eds.). Biodiversity Hotspots. Springer-Verlag, Berlin.

Prather, L.A., O. Alvarez-Fuentes, M.H. Mayfield, and C.J. Ferguson. 2004a. The decline in plant collecting in the United States: A threat to infrastructure of biodiversity studies. Syst. Bot. 29:15–28.

Prather, L.A., O. Alvarez-Fuentes, M.H. Mayfield, and C.J. Ferguson. 2004b. Implications of decline in plant collecting in the United States: A threat to infrastructure of biodiversity studies. Syst. Bot. 29: 216–220.

Preston, W. 1981. Vanishing Landscapes: Land and life in the Tulare Basin. Univ. of California Press, Berkeley.

Tewksbury, J. et. al. 2014. Natural history's place in science and society. BioScience 64: 300–310.

Thiers, B. [2014, continuously updated]. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. <http://sweetgum.nybg.org/ih/>
Tukey, J.W. 1977. Exploratory Data Analysis. Addison-Wesley, New York.
Welsh, S.L., N.D. Atwood, S. Goodrich, and L.C. Higgins. 2008. A Utah Flora (ed. 4). M.L. Bean Life Science Museum, Brigham Young Univ. Press, Provo, Utah
Whitfield, J. 2012. Rare specimens. Nature 484: 426–432.